09/744552
500 R[illegible] PCT/PTO 24 JAN 2001

Substitute Specification:

- - METHOD AND SYSTEM FOR DETERMINING MOVEMENT UNDERLYING A DIGITIZED IMAGE

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to digital video processing. In particular, the present invention relates to the determination of movement which underlies a digitized image.

Discussion of the Related Art

A method for determining a movement which underlies a digitized image is desxrcibed in, “A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera” by R. Mech, M. Wollborn, which appeared in Workshop on Image Analysis for Multimedia Interactive Services, Belgium, June 1997, as well as in an article by S. Colonnese et al., entitled “Adaptive Segmentation of Moving Object versus Background for Video Encoding” which appeared in Proceedings of SPIE Annual Symposium, Vol. 3164, San Diego, August 1997.

According to the Mech and Wollborn article, a global relative movement between a camera and a sequence of images taken by the camera is determined. Their method, which is used in the image stabilization of a camera, is based on a very inaccurate movement model which can describe only a tilting of the camera.

This disadvantage of a substantial inaccuracy in the determination of the global movement is also inherent to the method presented by Colonnese et al., which is used in the segmentation of the digitized image.

In order to achieve an improved accuracy, it is known to base the determination of a movement on a more complex movement model which is determined, with the aid of gradients in the digitized image, on the level of the pixels which are contained in the image, such as presented by S.S. Beauchemin, J.L. Barron in "The Computation of Optical Flow" ACM Computing Surveys, Vol. 27, No. 3, pages 366-433, September 1995. However, his method is complicated, and can therefore be carried out only with a substantial amount of computing time.

Furthermore, in the article entitled "Displacement Estimation by Hierarchical Blockmatching" by M. Bierlin, which appeared in SPIE, Vol. 1001, Visual Communications and Image Processing '88, pages 942 - 951, 1988, presents a method for so-called movement estimation for block-based image encoding. In this method, it is assumed that a digitized image has pixels which are grouped in image blocks of usually 8 x 8 pixels or 16 x 16 pixels. Furthermore, an image block is to be understood both as an image block of, for example 8 x 8 pixels or 16 x 16 pixels, and also a set of image blocks, for example a so-called macroblock, which contains 6 image blocks, of which, 4 image blocks hold brightness information and 2 image blocks hold color information.

Within the framework of a sequence of temporally succeeding images, for each image block the following method is carried out for an image to be coded for an image block in the image to be coded and a temporally preceding, already coded image: (1) an error value of an error dimension is formed for the image block, for

which a movement estimation is being carried out, in the temporally preceding image, starting from an image block which is located in the same relative position in the temporally preceding image, denoted below as a preceding image block, this being done, for example, by forming a sum over the absolute values of the differences of encoding information, assigned to the pixels, of the image block and the preceding image block. In this connection, encoding information is to be understood as brightness information (luminance value) and/or color information (chrominance value), which is respectively assigned to a pixel; (2) in a search space of prescribable size and shape about the initial position in the temporally preceding image, an error value of the error measure is formed in turn in each case in a region of the same size of an image block (preceding image block), displaced in each case by one or half a pixel; (3) this results in n^2 error values in a search space of size n * n pixels. That "displaced" preceding image block in the temporally preceding image is selected for which the error measure yields a minimum error value. It is assumed for this image block that this preceding image block corresponds best to the image block of the image to be coded for which the movement estimation is carried out; (4) the result of the movement estimation is a movement vector with which the displacement between the image block in the image to be coded and the selected image block in the temporally preceding image is described; (5) image data compression in the case of the block-based image encoding is achieved by virtue of the fact that only the movement vector and an error signal are coded; and (6) the movement estimation is carried out for each image block of an image.

However, the method described in the Bierlin article referred to above, cannot be used for a "global" movement estimation, which is the determination of

09455255 .012402

movement between a camera and the scene taken by the camera.

This is due to the heterogeneity of an image with a multiplicity of objects which are moving in different ways in the image. The application of the movement estimation to block-based image encoding, or to object-based image encoding, is discussed in ITU-T, International Telecommunication Union, Tele-communications Sector of ITU, Draft ITU-T Recommendation H.263, Video-Encoding for Low Bit-Rate Communication, 2nd May 1996.

The present invention is therefore based on solving the problem of determining and ascribing a movement which underlies a digitized image in a simple, fast and cost effective way, and can be used to improve the image segmentation method described by Colonnese et al., above.

The method for computer-aided determination of a movement which underlies a digitized image considers the digitized image contains pixels which are grouped into image blocks; a movement estimation is carried out for each image block, as a result of which a movement vector is determined for each image block, which movement vector is assigned to the respective image block; movement vectors are selected which are assigned to an image block which is situated in a prescribed region of the digitized image; parameters of a movement model are determined from the selected movement vectors; and the movement of the digitized image is described by the determined movement model.

The method and system for computer-aided determination of a movement which underlies a digitized image according to the present invention uses a processor which is set up in such a way that the digitized image contains pixels which are grouped into image blocks, a movement estimation is carried out for each image block

09445520.020401

image block, as a result of which a movement vector is determined for each image block, which movement vector is assigned to the respective image block, movement vectors are selected which are assigned to an image block which is situated in a prescribed region of the digitized image, parameters of a movement model are determined from the selected movement vectors, and the movement of the digitized image is described by the determined movement model.

The present invention provides an efficient, simple method and system, which can be carried out cost-effectively with a substantially reduced computing requirement. Furthermore, the present invention uses movement vectors which are determined by block-based image encoding, which itself is used to determine a global movement between a camera and a scene taken by the camera. However, when determining the movement, account is taken only of movement vectors which are assigned to image blocks situated in a prescribed region.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a method and system for determining movement underlying a digitized image wherein a prescribed region is formed by image blocks which are situated at a prescribed first distance from an edge of the digitized image and/or at a prescribed second distance from the middle of the digitized image.

It is another object of the present invention to provide a method and system for determining movement underlying a digitized image wherein movement vectors of image blocks which are situated at the edge of the image generally specify the

movement reliably.

It is a further object of the present invention to provide a method and system for determining movement underlying a digitized image wherein zooming and rotating of a camera can be specified reliably by movement vectors which are assigned to image blocks which are grouped in a region around the middle of the image.

It is an additional object of the present invention to provide a method and system for determining movement underlying a digitized image wherein the prescribed region clearly forms a "mask" in the form of a "perforated" rectangle inside the digitized image.

It is yet another object of the present invention to provide a method and system for determining movement underlying a digitized image involving the introduction of iterations to determine the movement model by modifying the "mask" after determining the parameters of the movement model and using this modified "mask" to recalculate the parameters of the movement model.

It is yet a further object of the present invention to provide a method and system for determining movement underlying a digitized image by forming the prescribed region by image blocks whose movement it was possible to estimate particularly reliably. This can be detected, for example, by virtue of the fact that the associated prediction error is below a prescribed threshold, or the variance of the prediction error in the search zone is above a threshold.

It is yet an additional object of the present invention to provide a method and system for determining movement underlying a digitized image wherein it is possible to use a "weighting mask" instead of the binary "mask", using blocks or their

movement vectors which are discretely selected for further calculation.

These and other objects and advantages of the present invention will become apparent upon careful review of the following detailed description of the preferred embodiments which is to be read in conjunction with review of the following drawing figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1	shows a block diagram according to the present invention;
Figure 2	shows a sketch of a coding and encoding of an image sequence according to the present invention;
Figure 3	shows an image encoding for global movement compensation according to the present invention;
Figures 4a - 4c	show processing of an image movement vector field according to the present invention; and
Figure 5	shows a flowchart according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Figure 1 shows, in block diagram form, the principle on which the global movement determination is based.

The parameters of the movement model 338 described below are calculated (step 103) starting from a movement vector field 101, the prescribed region or a weighting mask 102 and a weighting mask of reliability factors 106.

A movement vector field 101 is understood to be a set of all the determined movement vectors 330 relating to an image. The movement vector field 101 is illustrated (402) in Figure 4b by strokes which in each case describe a movement vector 330 for an image block. The movement vector field 402 is sketched on the digitized image 400. The image 400 comprises a moving object 403 in the form of a person, and an image background 404.

Figure 2 illustrates an arrangement which comprises two computers 202, 208 and a camera 201, image encoding, transmission of the image data and image decoding being illustrated.

A camera 201 is connected to a first computer 202 via a line 219. The camera 201 transmits taken images 204 to the first computer 202. The first computer 202 has a first processor 203, which is connected to an image store 205 via a bus 218. The method for image encoding is carried out with the aid of the first processor 203 of the first computer 202. Image data 206 encoded in this way are transmitted from the first computer 202 via a communication link 207, preferably a line or a radio path, to a second computer 208. The second computer 208 includes a second processor 209, which is connected to an image store 211 via a bus 210. A method for image decoding is carried out with the aid of the second processor 209.

Both the first computer 202 and the second computer 208 each have a display screen 212 and 213, respectively, on which the image data 204 are visualized. Input units, preferably a keyboard 214 and 215, respectively, and a computer mouse 216 and 217, respectively, are respectively provided for operating both the first computer 202 and the second computer 208.

The image data 204, which are transmitted to the first computer 202 by the camera 201 via the line 219 are data in the time domain, while the data 206, which are transmitted via the communication link 207 to the second computer 208 by the first computer 202 are image data in the spectral region. The decoded image data are illustrated on a display screen 220.

Figure 3 shows a sketch of an arrangement for carrying out a block-based image encoding method in accordance with the H.263 standard (see [5]).

A video data stream which is to be encoded and has temporally succeeding digitized images is fed to an image encoding unit 301. The digitized images are subdivided into macroblocks 302, each macroblock containing 16x16 pixels. The macroblock 302 comprises 4 image blocks 303, 304, 305 and 306, each image block containing 8 x 8 pixels to which luminance values (brightness values) are assigned. Each macroblock 302 further comprises two chrominance blocks 307 and 308 with chrominance values (color information, color saturation) assigned to the pixels.

The block of an image includes a luminance value (= brightness), a first chrominance value (= shade) and a second chrominance value (= color saturation). In this case, the luminance value, first chrominance value and second chrominance value are denoted as color values.

The image blocks are fed to a transformation encoding unit 309. In differential image encoding, values, to be encoded, of image blocks of temporally preceding images are subtracted from the image blocks currently to be encoded, and only the differential imaging information 310 is fed to the transformation encoding unit (Discrete Cosine Transformation, DCT) 309. For this purpose, the current

macroblock 302 is communicated via a connection 334 to a movement estimation unit 329. Spectral coefficients 311 are formed in the transformation encoding unit 309 for the image blocks or differential image blocks to be encoded, and are fed to a quantization unit 312.

Quantized spectral coefficients 313 are fed both to a scanning unit 314 and to an inverse quantization unit 315 in a return path. Entropy encoding is carried out on the scanned spectral coefficients 332 in an entropy encoding unit 316 provided therefor using a scanning method, for example a zigzag scanning method.

The entropy-encoded spectral coefficients are transmitted as encoded image data 317 to a decoder via a channel, preferably a line or a radio path.

Inverse quantization of the quantized spectral coefficients 313 is performed in the inverse quantization unit 315. Spectral coefficients 318 thus obtained are fed to an inverse transformation encoding unit 319 (Inverse Discrete Cosine Transformation, IDCT). Reconstructed encoding values (also differential encoding values) 320 are fed to an adder 321 in the differential image mode. The adder 321 also receives encoding values of an image block which result from a temporally preceding image after movement compensation which has already been carried out. Reconstructed image blocks 322 are formed with the aid of the adder 321 and stored in an image store 323.

Chrominance values 324 of the reconstructed image blocks 322 are fed from the image store 323 to a movement compensation unit 325. Interpolation in a specifically provided interpolation unit 327 is performed for brightness values 326. The interpolation is used to preferably double the number of brightness values contained in the respective image block. All brightness values 328 are fed both to

the movement compensation unit 325 and to the movement estimation unit 329. The movement estimation unit 329 also receives the image blocks of the particular macroblock (16x16 pixels) to be encoded, via the connection 334. The movement estimation is performed in the movement estimation unit 329 taking account of the interpolated brightness values ("movement estimation on a half-pixel basis").

The result of the movement estimation is a movement vector 330 which expresses a spatial displacement of the selected macroblock from the temporally preceding image to the macroblock 302 to be encoded.

Both brightness information and chrominance information relating to the macroblock determined by the movement estimation unit 329 are displaced by the movement vector 330 and subtracted from the encoding values of the macroblock 302, (see data path 231).

The way in which the movement estimation is performed is to determine for each image block for which a movement estimation is carried out an error E with respect to a zone of the same shape and size as the image block in a temporally preceding image, doing so, for example, in accordance with the following rule:

$$E = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| x_{i,j} - xd_{i,j} \right| \rightarrow \min \qquad \forall d \in S, \qquad (1)$$

- i, j denote respectively indices,

- n, m denote, respectively, a number (n) of pixels along a first direction x, and a number (m) of pixels along a second direction y, which are contained in the image block,

- $x_{i,j}$ denote respectively the encoding information which is assigned to a pixel at the relative position, denoted by the indices i, j, in the image block,

- $xd_{i,j}$ denote respectively the encoding information which is assigned to the respective pixel, denoted by i, j, in the zone of the temporally preceding image, displaced by a prescribable value d, and

- S denotes a searched space of prescribed shape and size in the temporally preceding image.

Calculation of the error E is carried out for each image block for different displacements within the search space S. That image block in the temporally preceding image whose error E is minimum is selected as most similar to the image block for which the movement estimation is carried out.

The result of the movement estimation is therefore yielded as the movement vector 330 with two movement vector components, a first movement vector component BV_x and a second movement vector component BV_y along the first direction x and the second direction y:

The movement vector 330 is assigned to the image block.

The image encoding unit from Figure 3 therefore supplies a movement vector 330 for all image blocks or macroimage blocks.

The movement vectors 330 are fed to a unit 335 for selecting or weighting the movement vectors 330. In the unit for selecting the movement vectors 335, those movement vectors 330 are selected or highly weighted which are assigned to image blocks which are located in a prescribed region 401 (compare Figure 4a). Furthermore, movement vectors which have been reliably (342) estimated are selected or highly weighted in the unit 335.

The selected movement vectors 336 are fed to a unit for determining the parameters of the movement model 337. The movement model in accordance with Figure 1, which is described below, is determined from the selected movement vectors in the unit for determining the parameters of the movement model 337. The determined movement model 338 is fed to a unit for compensating 339 the movement between the camera and the taken image. The movement is compensated in the unit for compensating 339 in accordance with a movement model described below, and so a movement-compensated image 340 which is less shaky is stored again, after processing in the unit for compensation 339, in the image store 323 in which the previously non-processed image whose movement is to be compensated is stored.

Figure 4a shows a prescribed region 401. The prescribed region 401 specifies a zone in which the image blocks must be situated so that the movement vectors which are assigned to these image blocks are selected.

The prescribed region 401 results from the fact that an edge region 405 which is formed by image blocks which are situated at a prescribed first distance of 406 from an edge 407 of the digitized image 400 [lacuna]. Image blocks are therefore not taken directly into account at the edge 407 of the image 400 when determining the parameters of the movement model 338. Furthermore, the prescribed region 401 is formed by image blocks which are situated at a prescribed second distance 408 from the middle 409 of the digitized image 400.

The prescribed region or the weighting mask is varied in an iterative method having the following steps to produce a new region of the following iteration (step 104).

For each image block in the prescribed region 401, a vector difference value VU is respectively determined, with the aid of which the difference of the determined movement model 338 with the movement vector 330 which is assigned to the respective image block is described. The vector difference value VU is formed, for example, in accordance with the following rule:

$$VU = \tfrac{1}{2}BV_X - MBV_X\tfrac{1}{2} + \tfrac{1}{2}BV_Y - MBV_Y\tfrac{1}{2}, \qquad (2)$$

MBV_x and MBV_y respectively denoting the components of a movement vector MBV calculated on the basis of the movement model.

The determination of the model-based movement vector is explained below in more detail.

In the case of the use of a binary mask, an image block is included in the new region of the further iteration when the respective vector differential value VU is smaller than a prescribable threshold value e. However, if the vector differential value VU is greater than the threshold value e the image block to which the respective movement vector is assigned is no longer taken into account in the new prescribed region. In the case of the use of a weighting mask, the weighting factors of the blocks are specified in the reverse ratio to that of the VU thereof.

As a result of this mode of procedure, those movement vectors which differ substantially from the movement vectors MBV calculated from the determined movement model are not taken into account, or are taken into account only slightly in calculating the parameters of the movement model in a further iteration.

After the new region or the new weighting mask has been formed, the movement vectors are used to assign the image blocks which are not included in the

new region, or a new set of parameters is determined for the movement model by making additional use of the weighting mask.

The method described above is carried out in a prescribable number of iterations or until a stop criterion, such as the undershooting of a number of eliminated blocks in an iteration step, for example, is fulfilled.

In this case, the new region is used in each case as the prescribed region or the new weighting mask in addition to the old movement vectors as input parameters of the next iteration. The determination of the global movement is carried out in such a way that parameters of a model for the global camera movement are determined.

A detailed derivation of the movement model is illustrated below in order to explain the movement model. It is assumed that a natural, three-dimensional scene is being projected by the camera onto a two-dimensional plane of projection. A projection of a point

$$\underline{p}_0 = (x_0, y_0, z_0)^T \tag{4}$$

is formed in accordance with the following rule:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \frac{F}{z_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \quad \wedge \quad z_0 >> F, \tag{5}$$

F describing a focal length and X,Y describing coordinates of the projected point $\underline{p}_0$ on the image plane.

If the camera is now moved, the projection rule is maintained in the coordinate system simultaneously moved synchronously with the camera, but the

coordinates of the object points must be transformed into this coordinate system. Since all the camera movements can be considered as an accumulation of rotation and translation, the transformation of the fixed coordinate system (x, y, z) into a simultaneously moved coordinate system $(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0)$ can be formulated in accordance with the following rule:

$$\begin{pmatrix} \tilde{x}_0 \\ \tilde{y}_0 \\ \tilde{z}_0 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}. \qquad (6)$$

Starting from rule (6) a change in image caused by camera movement is modeled in accordance with the following rule:

$$\begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} = \begin{pmatrix} C_F \cos(\varphi_z) - 1 & -C_F \sin(\varphi_z) \\ C_F \sin(\varphi_z) & C_F \cos(\varphi_z) - 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} t_X \\ t_Y \end{pmatrix}, \qquad (7)$$

DX, DY denoting a variation in the pixel coordinates caused in a time interval Dt in the case of the described camera movement, and j_z denoting the angle by which the camera has been rotated about a z-axis in this time interval Dt. A prescribed factor C_F denotes a change in focal length or a translation along the z axis.

The system of equations represented in rule (7) is nonlinear, for which reason the parameters of the system of equations cannot be determined directly.

Consequently, a simplified movement model is used for more rapid calculation, and in this case the camera movement in the plane of projection is used by a movement model with 6 parameters which are formed in accordance with the following rule:

$$\begin{pmatrix} \tilde{X}_0 \\ \tilde{Y}_0 \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix} \cdot \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \begin{pmatrix} t'_X \\ t'_Y \end{pmatrix}. \tag{8}$$

The system of equations produced therefrom with the data of the movement vector field is solved by means of linear regression, the complexity corresponding to inversion of a symmetrical 3 x 3 matrix.

After determination of the parameters r'_{11}, r'_{12}, r'_{21}, r'_{22}, t'_X and r'_y the parameters of rule (7) are approximated in accordance with the following rules:

$$\underline{T} = \underline{T}', \tag{9}$$

$$C_F = \sqrt{\left|\det\begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix}\right|}, \tag{10}$$

$$\rho_z = \arcsin \frac{1}{2}\left(r'_{21} - r'_{12}\right). \tag{11}$$

The movement which underlies an image relative to a camera which takes the image is compensated with the use of these parameters.

Figure 4c shows the movement vectors which are assigned to image blocks which are situated in the prescribed region 401. In this case, the prescribed region 401 is varied by an iteration (step 104) with respect to the prescribed region 401 from Figure 4a.

The method will be illustrated once again in terms of its individual method steps with the aid of Figure 5.

After the method has started (step 501), an image block or macroimage block is selected (step 502). A movement vector is determined (step 503) for the selected image block or macroimage block, and a check is made in a further step (step 504) as to whether all the image blocks or macroimage blocks of the image are processed.

If this is not the case, a further image block or macroimage block which has not yet been processed, is selected in a further step (step 505).

If, however, all the image blocks or macroimage blocks are processed, the movement vectors are selected which are assigned to an image block or a macroimage block which are situated in the prescribed region (step 506).

The parameters of the movement model are determined (step 507) from the selected movement vectors. If a further iteration is to be carried out, that is to say if the prescribed number of iterations has not yet been reached or the stop criterion is

09545024.120401

not yet fulfilled, a new region is determined in a further step (step 509), or the weighting mask of the next iteration is calculated as a function of the vector differential values VU (step 510). This is followed by compensating the movement of the image by using the determined movement model (step 508).

Some alternatives to the exemplary embodiment illustrated above are explained below:

The form of the region is fundamentally arbitrary and preferably dependent on prior knowledge of a scene. No use should be made in determining the movement model of those image regions of which it is known that these image regions differ clearly from the global movement.

The region should include only movement vectors of image regions which have proved to be reliable on the basis of the reliability values 342 of the movement estimation method.

In general, the movement estimation can be performed using any desired method, and is in no way limited to the principle of block matching. Thus, for example, movement estimation can also be performed using dynamic programming. Consequently, the type of movement estimation, and thus the way in which a movement vector is determined for an image block, are irrelevant to the present invention.

As an alternative to the approximate determination of the parameters of the system of equations (7), it is possible to linearize the sine terms and cosine terms in rule (7).

The following rule therefore results for small angles r_z

$$\begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} = \begin{pmatrix} C_F - 1 & -C_F\omega_z \\ C_F\omega_z & C_F - 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} t_X \\ t_Y \end{pmatrix} = \begin{pmatrix} R_1 & -R_2 \\ R_2 & R_1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} t_X \\ t_Y \end{pmatrix}. \tag{12}$$

Since the optimizations of the equations for DX and DY are not mutually independent, minimization is carried out with respect to the sum of the squares of the errors, that is to say in accordance with the following rule:

$$\sum_{\underline{V}} \left[\left(\Delta X_\eta - R_1 X_\eta + R_2 Y_\eta - t_x \right)^2 + \left(\Delta Y_\eta - R_2 X_\eta + R_1 Y_\eta - t_y \right)^2 \right] \rightarrow \min \tag{13}$$

Here, DX_h, DY_h denote the X- and Y-components, respectively, of the movement vector of the image block h at the position X_h, Y_h of the prescribed region $\underline{V}$ of the image.

In accordance with equation (12), R_1, R_2, t_x and t_y are the parameters of the movement model which are to be determined.

After the optimization method has been carried out, the associated model-based movement vector MBV (DX, DY) is determined on the basis of the determined system of equations (12) by substituting the X- and Y-components of the respective macroblock.

Instead of the abovenamed regions, it is also possible to make use of weighting masks A_X, A_Y which separately represent the reliability of the movement vectors, the a priori knowledge and the conclusions from the VU in the iterative

procedure for the X- and Y-components of the movement vectors when calculating the parameters of the movement model in accordance with the following optimization formulation:

$$\sum_{\underline{V}} \left[\begin{array}{l} \left(\alpha_{X_\eta} \cdot \left(\Delta X_\eta - R_1 X_\eta + R_2 Y_\eta - t_x\right)\right)^2 + \\ \left(\alpha_{Y_\eta} \cdot \left(\Delta Y_\eta - R_2 X_\eta - R_1 Y_\eta - t_x\right)\right)^2 \end{array} \right] \rightarrow \min$$

$$\alpha_{X_\eta} \in A_X, \alpha_{Y_\eta} \in A_Y. \tag{14}$$

A weighting mask A_x, A_y for the reliability of the movement vectors (105) can be calculated, for example, by calculating the values a_x, a_y for an image block in the following way in the case of block matching:

$$\alpha_X = \frac{1}{SAD_{match}} \cdot \sum_N \frac{|SAD_\eta - SAD_{match}|}{\left|x_\eta - x_{match}\right|}, \tag{15}$$

$$\alpha_y = \frac{1}{SAD_{match}} \cdot \sum_{N} \frac{|SAD_{\eta} - SAD_{match}|}{|y_{\eta} - y_{match}|}, \qquad (16)$$

SAD_h representing the sum of the pixel differences of a block for the h^{th} displacement (x_h, y_h) of the block matching, and SAD_{match} representing the same for the best, finally selected zone (x_{match}, y_{match}). N is the total number of search positions which have been investigated. If this value is calculated only taking account of the, for example, 16 best zones, the block matching can be carried out as a "spiral search" with the SAD of the worst of the 16 selected zones as stop criterion.

A further possibility of calculating a weighting mask $A_x = A_y = A$ for the reliability of the movement vectors is given by:

$$\alpha = \sum \frac{SAD - SAD_{match}}{N}, \qquad (17)$$

$a = a_x = a_y$ being the weighting factor of an image block or the movement vector thereof.

The present invention can be used, for example, to compensate a movement of a moving camera or also for the movement compensation of a camera which is integrated in a mobile communication unit, such as a video mobile phone.

According to the present invention, movement vectors which are determined during the block-based image encoding, can be used to determine a global movement between a camera and an image sequence taken by the camera. However, during determination of the movement account is taken only of movement

vectors which are assigned to image blocks which are situated in a prescribed region. The movement vectors of the image blocks are weighted in accordance with their reliability for the purpose of calculating the global movement.

Zooming and rotating of the video camera can be specified only unreliably by movement vectors which are assigned to image blocks which are grouped in a region around the middle of the image. In this case, the prescribed region clearly forms a "mask" in the form of a "perforated" rectangle inside the digitized image. Iterations are introduced to determine the movement model by modifying the "mask" after determining the parameters of the movement model. The modified "mask" is used to recalculate the parameters of the movement model. The "mask" can be modified by virtue of the fact that blocks whose movement vectors deviate from those of the movement model, and whose deviation exceeds a threshold value with reference to a prescribable distance measure, are eliminated from the prescribed region. The prescribed region is formed by image blocks whose movement can be estimated reliably, based upon an associated prediction error which is below a prescribed threshold, or the variance of the prediction error in the search zone is above a threshold. A "weighting mask" is used instead of the "binary mask" such that blocks or their movement vectors are weighted with factors. These can be different for the X-component and Y-component of the movement vector. The weightings feature in the calculation of the parameters of the movement model, and the determined movement can be used to compensate an actual movement of the arrangement with the aid of which an image is taken.

Although preferred embodiments of the present invention have been described herein, it is to be understood that the invention is not limited to these

embodiments and that various changes and modifications thereto may be made by persons having skill in the art to which the invention pertains, without departing from the scope or spirit of the invention, which is defined by the following claims. - -

4/PRB

09/744552
500 Rec'd PCT/PTO 2 4 JAN 2001

Description

Method and arrangement for determining a movement which underlies a digitized image

The invention relates to the determination of a movement which underlies a digitized image.

A method for determining a movement which underlies a digitized image is known from [1] and [2].

In the method from [1] a global relative movement between a camera and a sequence of images taken by the camera is determined. The method from [1], which is used in the image stabilization of a camera, is based on a very inaccurate movement model which can describe only a tilting of the camera.

This disadvantage of a substantial inaccuracy in the determination of the global movement is also inherent to the method from [2] which method is used in the segmentation of the digitized image.

In order to achieve an improved accuracy, it is known to base the determination of a movement on a more complex movement model which is determined, with the aid of gradients in the digitized image, on the level of the pixels which are contained in the image. However, this method is complicated, and can therefore be carried out only with a requirement for substantial computing time.

Furthermore [4] discloses a method for so-called movement estimation in a method for block-based image encoding. In this method, it is assumed that a digitized image has pixels which are grouped in image blocks of usually 8 * 8 pixels or 16 * 16 pixels.

Furthermore, an image block is to be understood both as an image block of, for example 8 * 8 pixels or 16 * 16 pixels, and also a set of image blocks, for example a so-called macroblock, which contains 6 image blocks (4 image blocks with brightness information, 2 image blocks with color information).

Within the framework of a sequence of temporally succeeding images, for each image block the following method is carried out for an image to be coded for an image block in the image to be coded and a temporally preceding, already coded image:

- An error value of an error dimension is formed for the image block, for which a movement estimation is being carried out, in the temporally preceding image, starting from an image block which is located in the same relative position in the temporally preceding image, denoted below as a preceding image block, this being done, for example, by forming a sum over the absolute values of the differences of encoding information, assigned to the pixels, of the image block and the preceding image block.

In this connection, encoding information is to be understood as brightness information (luminance value) and/or color information (chrominance value), which is respectively assigned to a pixel.

- In a search space of prescribable size and shape about the initial position in the temporally preceding image, an error value of the error measure is formed in turn in each case in a region of the same size of an image block (preceding image block), displaced in each case by one or half a pixel.
- This results in n^2 error values in a search space of size n * n pixels. That "displaced" preceding image block in the temporally preceding image is selected for

which the error measure yields a minimum error value. It is assumed for this image block that this preceding image block corresponds best to the image block of the image to be coded for which the movement estimation is carried out.

- The result of the movement estimation is a movement vector with which the displacement between the image block in the image to be coded and the selected image block in the temporally preceding image is described.

- Image data compression in the case of the block-based image encoding is achieved by virtue of the fact that only the movement vector and an error signal are coded.

- The movement estimation is carried out for each image block of an image.

However, the method described in [4] cannot be used for a "global" movement estimation, that is to say determination of the movement between a camera and the scene taken by the camera.

This is ascribed, in particular, to the heterogeneity of an image with a multiplicity of objects which are moving in different ways in the image.

The application of the movement estimation to block-based image encoding, or else to object-based image encoding is known from [5] and [6].

The invention is therefore based on the problem of determining and ascribing a movement which underlies a digitized image in a simple, fast and cost effective way.

The problem is solved by means of the method in accordance with patent claim 1, and by means of the arrangement in accordance with patent claim 10.

The method for computer-aided determination of a movement which underlies a digitized image comprises the following steps:

- the digitized image contains pixels which are grouped into image blocks,
- a movement estimation is carried out for each image block, as a result of which a movement vector is determined for each image block, which movement vector is assigned to the respective image block,
- movement vectors are selected which are assigned to an image block which is situated in a prescribed region of the digitized image,
- parameters of a movement model are determined from the selected movement vectors, and
- the movement of the digitized image is described by the determined movement model.

The arrangement for computer-aided determination of a movement which underlies a digitized image has a processor which is set up in such a way that the following steps can be carried out:

- the digitized image contains pixels which are grouped into image blocks,
- a movement estimation is carried out for each image block, as a result of which a movement vector is determined for each image block, which movement vector is assigned to the respective image block,
- movement vectors are selected which are assigned to an image block which is situated in a prescribed region of the digitized image,
- parameters of a movement model are determined from the selected movement vectors, and
- the movement of the digitized image is described by the determined movement model.

The method provides an efficient, simple method, which can therefore be carried out cost-effectively with a substantially lesser computing requirement, and an arrangement which can therefore be implemented cost-effectively.

The invention is to be seen clearly in that movement vectors which are determined in any case with the block-based image encoding are used to determine a global movement between a camera and a scene taken by the camera.

However, when determining the movement account is taken only of movement vectors which are assigned to image blocks which are situated in a prescribed region.

Advantageous developments of the invention follow from the dependent claims.

In a development of the invention, it is advantageous that the prescribed region is formed by image blocks which are situated at a prescribed first distance from an edge of the digitized image and/or at a prescribed second distance from the middle of the digitized image.

This development is based on the finding that movement vectors of image blocks which are situated at the edge of the image generally specify the actual movement only unreliably. Furthermore, zooming and rotating of a camera can be specified only unreliably by movement vectors which are assigned to image blocks which are grouped in a region around the middle of the image.
In this case, the prescribed region clearly forms a "mask" in the form of a "perforated" rectangle inside the digitized image.

A further development consists in introducing iterations in determining the movement model by modifying the "mask" after determining the parameters of the movement model and using this modified "mask" to recalculate the parameters of the movement model. The "mask" can be modified in this case, for example, by virtue of the fact that blocks whose movement vectors deviate from those of the movement model, and this deviation exceeds a threshold value with reference to a prescribable distance measure, are eliminated from the prescribed region.

A further refinement consists in forming the prescribed region by image blocks whose movement it was possible to estimate particularly reliably. This can be detected, for example, by virtue of the fact that the associated prediction error is below a prescribed threshold, or the variance of the prediction error in the search zone is above a threshold.

Furthermore, it is possible to use a "weighting mask" instead of the binary "mask" described in the foregoing paragraphs. In this case, it is not, as previously described, blocks or their movement vectors which are discretely selected for further calculation, but the blocks or their movement vectors are weighted with factors. These can be different for the X-component and Y-component of the movement vector. These weightings feature in the calculation of the parameters of the movement model.

The determined movement can be used to compensate an actual movement of the arrangement with the aid of which an image is taken.
The invention can be used to compensate a camera movement or also to compensate a movement of a mobile communication device which includes the camera.

An exemplary embodiment of the invention is illustrated in the drawings and explained in more detail below.

In the drawing:

Figure 1 shows a block diagram in which the principle of the exemplary embodiment is illustrated pictorially;

Figure 2 shows a sketch of an arrangement with a camera and an encoding unit for encoding the image sequence taken with the camera, and an arrangement for decoding the encoded image sequence;

Figure 3 shows a detailed sketch of the arrangement for image encoding and for global movement compensation;

Figures 4a to c respectively show an image in which a movement vector field is determined for the image relative to a temporally preceding image with a prescribed region (Figure 1a) from which in each case the movement vectors are determined for forming parameters of a movement model, an image with all the movement vectors (Figure 1b) and an image with movement vectors after iteration of the method with the prescribed region illustrated in Figure 1a (Figure 1c);

Figure 5 shows a flowchart in which the method steps of the exemplary embodiment are illustrated.

Figure 2 illustrates an arrangement which comprises two computers 202, 208 and a camera 201, image encoding, transmission of the image data and image decoding being illustrated.

A camera 201 is connected to a first computer 202 via a line 219. The camera 201 transmits taken images 204 to the first computer 202. The first computer 202 has a first processor 203, which is connected to an image store 205 via a bus 218. The method for image encoding is carried out with the aid of the first processor 203 of the first computer 202. Image data 206 encoded in this way are transmitted from the first computer 202 via a communication link 207, preferably a line or a radio path, to a second computer 208. The second computer 208 includes a second processor 209, which is connected to an image store 211 via a bus 210. A method for image decoding is carried out with the aid of the second processor 209.

Both the first computer 202 and the second computer 208 each have a display screen 212 and 213, respectively, on which the image data 204 are visualized. Input units, preferably a keyboard 214 and 215, respectively, and a computer mouse 216 and 217, respectively, are respectively provided for operating both the first computer 202 and the second computer 208.

The image data 204, which are transmitted to the first computer 202 by the camera 201 via the line 219 are data in the time domain, while the data 206, which are transmitted via the communication link 207 to the second computer 208 by the first computer 202 are image data in the spectral region.

The decoded image data are illustrated on a display screen 220.

Figure 3 shows a sketch of an arrangement for carrying out a block-based image encoding method in accordance with the H.263 standard (see [5]).

A video data stream which is to be encoded and has temporally succeeding digitized images is fed to an

image encoding unit 301. The digitized images are subdivided into macroblocks 302, each macroblock containing 16x16 pixels. The macroblock 302 comprises 4 image blocks 303, 304, 305 and 306, each image block containing 8x8 pixels to which luminance values (brightness values) are assigned. Each macroblock 302 further comprises two chrominance blocks 307 and 308 with chrominance values (color information, color saturation) assigned to the pixels.

The block of an image includes a luminance value (= brightness), a first chrominance value (= shade) and a second chrominance value (= color saturation). In this case, the luminance value, first chrominance value and second chrominance value are denoted as color values.

The image blocks are fed to a transformation encoding unit 309. In differential image encoding, values, to be encoded, of image blocks of temporally preceding images are subtracted from the image blocks currently to be encoded, and only the differential imaging information 310 is fed to the transformation encoding unit (Discrete Cosine Transformation, DCT) 309. For this purpose, the current macroblock 302 is communicated via a connection 334 to a movement estimation unit 329. Spectral coefficients 311 are formed in the transformation encoding unit 309 for the image blocks or differential image blocks to be encoded, and are fed to a quantization unit 312.

Quantized spectral coefficients 313 are fed both to a scanning unit 314 and to an inverse quantization unit 315 in a return path. Entropy encoding is carried out on the scanned spectral coefficients 332 in an entropy encoding unit 316 provided therefor using a scanning method, for example a zigzag scanning method. The entropy-encoded spectral coefficients are

transmitted as encoded image data 317 to a decoder via a channel, preferably a line or a radio path.

Inverse quantization of the quantized spectral coefficients 313 is performed in the inverse quantization unit 315. Spectral coefficients 318 thus obtained are fed to an inverse transformation encoding unit 319 (Inverse Discrete Cosine Transformation, IDCT). Reconstructed encoding values (also differential encoding values) 320 are fed to an adder 321 in the differential image mode. The adder 321 also receives encoding values of an image block which result from a temporally preceding image after movement compensation which has already been carried out. Reconstructed image blocks 322 are formed with the aid of the adder 321 and stored in an image store 323.

Chrominance values 324 of the reconstructed image blocks 322 are fed from the image store 323 to a movement compensation unit 325. Interpolation in a specifically provided interpolation unit 327 is performed for brightness values 326. The interpolation is used to preferably double the number of brightness values contained in the respective image block. All brightness values 328 are fed both to the movement compensation unit 325 and to the movement estimation unit 329. The movement estimation unit 329 also receives the image blocks of the particular macroblock (16x16 pixels) to be encoded, via the connection 334. The movement estimation is performed in the movement estimation unit 329 taking account of the interpolated brightness values ("movement estimation on a half-pixel basis").

The result of the movement estimation is a movement vector 330 which expresses a spatial displacement of the selected

macroblock from the temporally preceding image to the macroblock 302 to be encoded.

Both brightness information and chrominance information relating to the macroblock determined by the movement estimation unit 329 are displaced by the movement vector 330 and subtracted from the encoding values of the macroblock 302 (see data path 231).

The way in which the movement estimation is performed is to determine for each image block for which a movement estimation is carried out an error E with respect to a zone of the same shape and size as the image block in a temporally preceding image, doing so, for example, in accordance with the following rule:

$$E = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| x_{i,j} - xd_{i,j} \right| \rightarrow \min \qquad \forall d \in S, \qquad (1)$$

- i, j denote respectively indices,
- n, m denote, respectively, a number (n) of pixels along a first direction x, and a number (m) of pixels along a second direction y, which are contained in the image block,
- $x_{i,j}$ denote respectively the encoding information which is assigned to a pixel at the relative position, denoted by the indices i, j, in the image block,
- $xd_{i,j}$ denote respectively the encoding information which is assigned to the respective pixel, denoted by i, j, in the zone of the temporally preceding image, displaced by a prescribable value d, and
- S denotes a searched space of prescribed shape and size in the temporally preceding image.

The calculation of the error E is carried out for each image block for different displacements within the search space S. That image block in the temporally preceding image whose error E is minimum is selected as

most similar to the image block for which the movement estimation is carried out.

The result of the movement estimation is therefore yielded as the movement vector 330 with two movement vector components, a first movement vector component BV_x and a second movement vector component BV_y along the first direction x and the second direction y:

$$BV = \begin{pmatrix} BV_x \\ BV_y \end{pmatrix}.$$

The movement vector 330 is assigned to the image block.

The image encoding unit from Figure 3 therefore supplies a movement vector 330 for all image blocks or macroimage blocks.

The movement vectors 330 are fed to a unit 335 for selecting or weighting the movement vectors 330. In the unit for selecting the movement vectors 335, those movement vectors 330 are selected or highly weighted which are assigned to image blocks which are located in a prescribed region 401 (compare Figure 4a). Furthermore, movement vectors which have been reliably (342) estimated are selected or highly weighted in the unit 335.

The selected movement vectors 336 are fed to a unit for determining the parameters of the movement model 337. The movement model in accordance with Figure 1, whichis described below, is determined from the selected movement vectors in the unit for determining the parameters of the movement model 337.

09744553 .012401

The determined movement model 338 is fed to a unit for compensating 339 the movement between the camera and the taken image. The movement is compensated in the unit for compensating 339 in accordance with a movement model described below, and so a movement-compensated image 340 which is less shaky is stored again, after processing in the unit for compensation 339, in the image store 323 in which the previously non-processed image whose movement is to be compensated is stored.

Figure 1 shows in the form of a block diagram the principle on which the global movement determination is based.

The parameters of the movement model 338 described below are calculated (step 103) starting from a movement vector field 101, the prescribed region or a weighting mask 102 and a weighting mask of reliability factors 106.

A movement vector field 101 is understood to be a set of all the determined movement vectors 330 relating to an image. The movement vector field 101 is illustrated (402) in **Figure 4b** by strokes which in each case describe a movement vector 330 for an image block. The movement vector field 402 is sketched on the digitized image 400. The image 400 comprises a moving object 403 in the form of a person, and an image background 404.

Figure 4a shows a prescribed region 401. The prescribed region 401 specifies a zone in which the image blocks must be situated so that the movement vectors which are assigned to these image blocks are selected.

The prescribed region 401 results from the fact that an edge region 405 which is formed by image blocks which are situated at a prescribed first distance of 406 froman edge 407 of the digitized image 400 [lacuna].

Image blocks are therefore not taken directly into account at the edge 407 of the image 400 when determining the parameters of the movement model 338. Furthermore, the prescribed region 401 is formed by image blocks which are situated at a prescribed second distance 408 from the middle 409 of the digitized image 400.

The prescribed region or the weighting mask is varied in an iterative method having the following steps to produce a new region of the following iteration (step 104).

For each image block in the prescribed region 401, a vector difference value VU is respectively determined, with the aid of which the difference of the determined movement model 338 with the movement vector 330 which is assigned to the respective image block is described. The vector difference value VU is formed, for example, in accordance with the following rule:

$$VU = \left|BV_X - MBV_X\right| + \left|BV_Y - MBV_Y\right|, \qquad (2)$$

MBV_x and MBV_y respectively denoting the components of a movement vector MBV calculated on the basis of the movement model.

The determination of the model-based movement vector is explained below in more detail.

In the case of the use of a binary mask, an image block is included in the new region of the further iteration when the respective vector differential value VU is smaller than a prescribable threshold value ε. However, if the vector differential value VU is greater than the threshold value ε the image block to which the respective movement vector is assigned is no longer taken into account in the new prescribed region.

In the case of the use of a weighting mask, the weighting factors of the blocks are specified in the reverse ratio to that of the VU thereof.

As a result of this mode of procedure, those movement vectors which differ substantially from the movement vectors MBV calculated from the determined movement model are not taken into account, or are taken into account only slightly in calculating the parameters of the movement model in a further iteration.

After the new region or the new weighting mask has been formed, the movement vectors are used to assign the image blocks which are not included in the new region, or a new set of parameters is determined for the movement model by making additional use of the weighting mask.

The method described above is carried out in a prescribable number of iterations or until a stop criterion, such as the undershooting of a number of eliminated blocks in an iteration step, for example, is fulfilled.

In this case, the new region is used in each case as the prescribed region or the new weighting mask in addition to the old movement vectors as input parameters of the next iteration.

The determination of the global movement is carried out in such a way that parameters of a model for the global camera movement are determined.

A detailed derivation of the movement model is illustrated below in order to explain the movement model:It is assumed that a natural, three-dimensional scene is being projected by the camera onto a two-

dimensional plane of projection. A projection of a point

$$\underline{p}_0 = (x_0, y_0, z_0)^T \tag{4}$$

is formed in accordance with the following rule:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \frac{F}{z_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \quad \wedge \quad z_0 >> F, \tag{5}$$

F describing a focal length and X,Y describing coordinates of the projected point $\underline{p}_0$ on the image plane.

If the camera is now moved, the projection rule is maintained in the coordinate system simultaneously moved synchronously with the camera, but the coordinates of the object points must be transformed into this coordinate system. Since all the camera movements can be considered as an accumulation of rotation and translation, the transformation of the fixed coordinate system (x, y, z) into a simultaneously moved coordinate system $(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0)$ can be formulated in accordance with the following rule:

$$\begin{pmatrix} \tilde{x}_0 \\ \tilde{y}_0 \\ \tilde{z}_0 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}. \tag{6}$$

Starting from rule (6) a change in image caused by camera movement is modeled in accordance with the following rule:

$$\begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} = \begin{pmatrix} C_F \cos(\varphi_z) - 1 & - C_F \sin(\varphi_z) \\ C_F \sin(\varphi_z) & C_F \cos(\varphi_z) - 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} t_X \\ t_Y \end{pmatrix}, \tag{7}$$

ΔX, ΔY denoting a variation in the pixel coordinates caused in a time interval Δt in the case of the

described camera movement, and φ_z denoting the angle by which the camera has been rotated about a z-axis in this time interval Δt. A prescribed factor C_F denotes a change in focal length or a translation along the z axis.

The system of equations represented in rule (7) is nonlinear, for which reason the parameters of the system of equations cannot be determined directly.

Consequently, a simplified movement model is used for more rapid calculation, and in this case the camera movement in the plane of projection is used by a movement model with 6 parameters which are formed in accordance with the following rule:

$$\begin{pmatrix} \tilde{X}_0 \\ \tilde{Y}_0 \end{pmatrix} = \begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix} \cdot \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \begin{pmatrix} t'_X \\ t'_Y \end{pmatrix}. \tag{8}$$

The system of equations produced therefrom with the data of the movement vector field is solved by means of linear regression, the complexity corresponding to inversion of a symmetrical 3 * 3 matrix.

After determination of the parameters r'_{11}, r'_{12}, r'_{21}, r'_{22}, t'_x and r'_y the parameters of rule (7) are approximated in accordance with the following rules:

$$\underline{T} = \underline{T}', \tag{9}$$

$$C_F = \sqrt{\left| \det \begin{pmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{pmatrix} \right|}, \tag{10}$$

$$\rho_z = \arcsin \frac{1}{2} \left(r'_{21} - r'_{12} \right). \tag{11}$$

The movement which underlies an image relative to a camera which takes the image is compensated with the use of these parameters.

Figure 4c shows the movement vectors which are assigned to image blocks which are situated in the prescribed region 401. In this case, the prescribed region 401 is varied by an iteration (step 104) with respect to the prescribed region 401 from Figure 4a.

The method will be illustrated once again in terms of its individual method steps with the aid of **Figure 5**:

After the method has started (step 501), an image block or macroimage block is selected (step 502). A movement vector is determined (step 503) for the selected image block or macroimage block, and a check is made in a further step (step 504) as to whether all the image blocks or macroimage blocks of the image are processed.

If this is not the case, a further image block or macroimage block which has not yet been processed, is selected in a further step (step 505).

If, however, all the image blocks or macroimage blocks are processed, the movement vectors are selected which are assigned to an image block or a macroimage block which are situated in the prescribed region (step 506).

The parameters of the movement model are determined (step 507) from the selected movement vectors. If a further iteration is to be carried out, that is to say if the prescribed number of iterations has not yet been reached or the stop criterion is not yet fulfilled, a new region is determined in a further step (step 509), or the weighting mask of the next iteration is calculated as a function of the vector differential values VU (step 510).

This is followed by compensating the movement of the image by using the determined movement model (step 508).

Some alternatives to the exemplary embodiment illustrated above are explained below:

The form of the region is fundamentally arbitrary and preferably dependent on prior knowledge of a scene. No use should be made in determining the movement model of those image regions of which it is known that these image regions differ clearly from the global movement.

The region should include only movement vectors of image regions which have proved to be reliable on the basis of the reliability values 342 of the movement estimation method.

In general, the movement estimation can be performed using any desired method, and is in no way limited to the principle of block matching. Thus, for example, movement estimation can also be performed using dynamic programming.

Consequently, the type of movement estimation, and thus the way in which a movement vector is determined for an image block, are irrelevant to the invention.

As an alternative to the approximate determination of the parameters of the system of equations (7), it is possible to linearize the sine terms and cosine terms in rule (7).

The following rule therefore results for small angles ρ_z

$$\begin{pmatrix}\Delta X\\ \Delta Y\end{pmatrix} = \begin{pmatrix}C_F - 1 & -C_F\omega_z\\ C_F\omega_z & C_F - 1\end{pmatrix}\cdot\begin{pmatrix}X\\ Y\end{pmatrix} + \begin{pmatrix}t_X\\ t_Y\end{pmatrix} = \begin{pmatrix}R_1 & -R_2\\ R_2 & R_1\end{pmatrix}\cdot\begin{pmatrix}X\\ Y\end{pmatrix} + \begin{pmatrix}t_X\\ t_Y\end{pmatrix}. \tag{12}$$

Since the optimizations of the equations for ΔX and ΔY are not mutually independent, minimization is carried out with respect to the sum of the squares of the errors, that is to say in accordance with the following rule:

$$\sum_{\underline{V}}\left[\left(\Delta X_\eta - R_1 X_\eta + R_2 Y_\eta - t_x\right)^2 + \left(\Delta Y_\eta - R_2 X_\eta + R_1 Y_\eta - t_y\right)^2\right] \rightarrow \min \tag{13}$$

Here, ΔX_η, ΔY_η denote the X- and Y-components, respectively, of the movement vector of the image block η at the position X_η, Y_η of the prescribed region $\underline{V}$ of the image.

In accordance with equation (12), R_1, R_2, t_x and t_y are the parameters of the movement model which are to be determined.

After the optimization method has been carried out, the associated model-based movement vector MBV (ΔX, ΔY) is determined on the basis of the determined system of equations (12) by substituting the X- and Y-components of the respective macroblock.

Instead of the abovenamed regions, it is also possible to make use of weighting masks A_x, A_Y which separately represent the reliability of the movement vectors, the a priori knowledge and the conclusions from the VU in the iterative procedure for the X- and Y-components of the movement vectors when calculating the parameters of the movement model in accordance with the following optimization formulation:

$$\sum_{\underline{V}} \left[\begin{array}{l} \left(\alpha_{X_\eta} \cdot \left(\Delta X_\eta - R_1 X_\eta + R_2 Y_\eta - t_x\right)\right)^2 + \\ \left(\alpha_{Y_\eta} \cdot \left(\Delta Y_\eta - R_2 X_\eta - R_1 Y_\eta - t_x\right)\right)^2 \end{array} \right] \rightarrow \min$$

$$\alpha_{X_\eta} \in A_X, \alpha_{Y_\eta} \in A_Y. \tag{14}$$

A weighting mask A_x, A_y for the reliability of the movement vectors (105) can be calculated, for example, by calculating the values α_x, α_y for an image block in the following way in the case of block matching:

$$\alpha_x = \frac{1}{SAD_{match}} \cdot \sum_N \frac{|SAD_\eta - SAD_{match}|}{|x_\eta - x_{match}|}, \tag{15}$$

$$\alpha_y = \frac{1}{SAD_{match}} \cdot \sum_N \frac{|SAD_\eta - SAD_{match}|}{|y_\eta - y_{match}|}, \tag{16}$$

SAD_η representing the sum of the pixel differences of a block for the η^{th} displacement (x_η, y_η) of the block matching, and SAD_{match} representing the same for the best, finally selected zone (x_{match}, y_{match}). N is the total number of search positions which have been investigated. If this value is calculated only taking account of the, for example, 16 best zones, the block matching can be carried out as a "spiral search" with the SAD of the worst of the 16 selected zones as stop criterion.

A further possibility of calculating a weighting mask $A_x = A_y = A$ for the reliability of the movement vectors is given by:

$$\alpha = \sum \frac{SAD - SAD_{match}}{N}, \qquad (17)$$

$\alpha = \alpha_x = \alpha_y$ being the weighting factor of an image block or the movement vector thereof.

The invention can be used, for example, to compensate a movement of a moving camera or
also for the movement compensation of a camera which is integrated in a mobile communication unit (video mobile phone).

The invention can in addition be used for image segmentation as described in [2].

The invention is to be seen vividly in that movement vectors which are determined in any case during the block-based image encoding are used to determine a global movement between a camera and an image sequence taken by the camera.

However, during determination of the movement account is taken only of movement vectors which are assigned to image blocks which are situated in a prescribed region.

The movement vectors of the image blocks are weighted in accordance with their reliability for the purpose of calculating the global movement.
The following publications are cited in this document:

[1] R. Mech, M. Wollborn, A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera, Workshop on Image Analysis for Multimedia Interactive Services, Belgium, June 1997

[2] S. Colonnese et al., Adaptive Segmentation of Moving Object versus Background for Video